



UNITED STATES PATENT AND TRADEMARK OFFICE

CP
UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/848,982	05/03/2001	Ted E. Dunning	22227-05479	8782
758	7590	01/14/2005	EXAMINER	
FENWICK & WEST LLP SILICON VALLEY CENTER 801 CALIFORNIA STREET MOUNTAIN VIEW, CA 94041			WONG, LESLIE	
			ART UNIT	PAPER NUMBER
			2167	

DATE MAILED: 01/14/2005

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary	Application No.	Applicant(s)	
	09/848,982	DUNNING ET AL.	
	Examiner	Art Unit	
	Leslie Wong	2167	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) Responsive to communication(s) filed on 20 September 2004.
- 2a) This action is FINAL. 2b) This action is non-final.
- 3) Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) Claim(s) 1-35 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) Claim(s) _____ is/are allowed.
- 6) Claim(s) 1-35 is/are rejected.
- 7) Claim(s) _____ is/are objected to.
- 8) Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) The specification is objected to by the Examiner.
- 10) The drawing(s) filed on 03 May 2001 is/are: a) accepted or b) objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) All b) Some * c) None of:
 1. Certified copies of the priority documents have been received.
 2. Certified copies of the priority documents have been received in Application No. _____.
 3. Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) <input type="checkbox"/> Notice of References Cited (PTO-892)	4) <input type="checkbox"/> Interview Summary (PTO-413)
2) <input type="checkbox"/> Notice of Draftsperson's Patent Drawing Review (PTO-948)	Paper No(s)/Mail Date. _____
3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____	5) <input type="checkbox"/> Notice of Informal Patent Application (PTO-152)
	6) <input type="checkbox"/> Other: _____

DETAILED ACTION

Response to Amendment

1. Receipt of Applicant's Amendment, filed 20 September 2004, is acknowledged.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was not commonly owned at the time a later invention was made in order for the examiner to consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g) prior art under 35 U.S.C. 103(a).

3. Claims 1-35 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Damashek** (U.S. Patent 5,418,951) in view of **Ortega et al.** (U.S. Patent Application 20020152204A1).

Regarding claims 1, 12, 23, and 34, **Damashek teaches a computer-implemented method, system, and computer-readable medium for performing text equivalencing from a query string of characters comprising:**

a). **'modifying the query string using a predetermined set of heuristics'** as reducing multiple spaces to a single space within a string of characters and the strings of characters may also be eliminated or replaced by a user-defined character or strings of characters (col. 4, line 64 – col. 5, line 5; col. 8, line 64 col. 9, line 2);

b). **'comparing the modified query string with at least one known string of characters in a corpus in order to locate a match'** as comparing the scores for the n-grams strings between the unidentified document and the reference documents to determine the degree of similarity between the strings of the two documents (col. 5, lines 54-67; col. 4, lines 10-60);

c). **'responsive to not finding an exact match, performing the steps of:**

1). **'forming a plurality of sub-strings of characters from the query string, the sub-strings having varying lengths'** as parsing text which is written in an unidentified language into n-grams. N-grams (i.e. sub-strings) are consecutive runs of n characters where n is any positive integer greater than zero (col. 4, lines 49-56; col. 5, lines 24-30; col. 3, lines 21-24; col. 4, lines 24-27); and

2). **'using an information retrieval technique on the sub-strings formed from the query string to identify a known string of characters'**

equivalent to the query string' as enumerating the n-grams contained in the unidentified document and comparing the result of that operation with the enumerated n-grams found in a reference document (col. 3, lines 22-34 and col. 4, lines 10-60).

b). **Damashek** does not explicitly teach a step of performing a character-by-character comparison of the strings.

Ortega et al., however, teaches a step of '**performing a character-by-character comparison of the query string**' as comparing a non-matching term to the list of related terms one-by-one using an anagram-type function which compares two character-strings and returns a numerical similarity score (¶s 0021, 0033, 0057-0064).

It would have been obvious to one of ordinary skill in the art at the time of the invention was made to combine the teachings of the cited references because **Ortega's** teaching would have allowed **Damashek's** to facilitate processing and increasing the efficiency of a search query by invoking the spelling correction process to attempt to correct the non-matching term(s) and comparing a non-matching term of a search criteria with data in the correlation table to identify any possible replacements (¶ 0010) as suggested by **Ortega et al.** at ¶s 0054 and 0066.

Regarding claims 2, 13, 24, and 35, **Damashek** further teaches a step wherein the information retrieval technique further comprises:

a). weighting the sub-strings (col. 5, lines 31);

- b). scoring the known string of characters (col. 8, lines 51-56); and
- c). retrieving information associated with a known string having the highest score (col. 9, lines 64-66).

Regarding claims 3, 14, and 25, **Damashek** further teaches a step comprising, responsive to the highest score being greater than a first threshold, automatically accepting the known string having the highest score as an exact match (col. 8, lines 51-63).

Regarding claims 4, 15, and 26, **Damashek** further teaches a step comprising, responsive to the highest score being less than a second threshold and greater than a first threshold, presenting the known string having the highest score to a user for manual confirmation (col. 9, lines 12-14; col. 10. 45-49).

Regarding claims 5, 16, and 27, **Damashek** further teaches a step comprising, responsive to the highest score being less than a second threshold and greater than a third threshold, presenting the known string having he highest score to a user to select the equivalent string of characters (col. 9, lines 12-14; col. 10. 45-49).

Regarding claims 6, 17, and 28, **Damashek** further teaches a step forming a plurality of sub-strings of characters comprises successively extending sub-strings

based on frequency of occurrence in the modified query string (col. 3, lines 21-24; col. 4, lines 24-27).

Regarding claims 7, 18, and 29, **Damashek** further teaches a step wherein the query string is selected from the group consisting of a song title, a song artist, an album name, a book title, an author's name, a book publisher, a genetic sequence, and a computer program (col. 9, lines 35-37).

Regarding claims 8, 19, and 30, **Damashek** further teaches a step wherein the predetermined set of heuristics comprises removing whitespace from the query string (col. 4, line 64 – col. 5, line 5).

Regarding claims 9, 20, and 31, **Damashek** further teaches a step wherein the predetermined set of heuristics comprises removing a portion of the query string (col. 8, line 64 – col. 9, line 10).

Regarding claims 10, 21, and 32, **Damashek** further teaches a step wherein the predetermined set of heuristics comprises replacing a symbol in the query string with an alternate representation for the symbol (col. 4, line 64 – col. 5, line 5).

Regarding claims 11, 22, and 33, **Damashek** further teaches a step wherein storing a database entry indicating (i.e., similarity score) that the query string is an equivalent of the identified known string (col. 8, lines 51-56).

Response to Argument

4. Applicants' arguments filed 20 September 2004 have been fully considered but they are not persuasive.

Applicants argue that Damashek and Ortega are directed at completely different problems: Damashek presents a method for identifying the topic or language of a document, while Ortega describes techniques for predicting correct spellings in multiple-term queries. There is no hint or suggestion in either reference of combining the references in the manner suggested by the Examiner. In response to applicant's argument that there is no suggestion to combine the references, the examiner recognizes that obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention where there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art. See *In re Fine*, 837 F.2d 1071, 5 USPQ2d 1596 (Fed. Cir. 1988) and *In re Jones*, 958 F.2d 347, 21 USPQ2d 1941 (Fed. Cir. 1992). In this case Applicants' invention concerns with finding equivalent text. Equivalent texts are two pieces of text that are intended to be exactly the same, at least one of which contains a misspelling or typographical error or a

difference in representation such as using a symbolic representation of a word or a different word order. Finding equivalent text is desirable in identifying a song or book by its textual identifiers such as title, artist or author, album name or publisher name etc... (Applicants' Specification paragraph 0003 and 0007). Applicants' invention modifies the input string to eliminate common mistakes such as eliminating or adding whitespace, adding or deleting the word "the", changing "&" to "and", and a few common misspellings are typographical error (Applicants' Specification paragraph 0008). The modified strings are then compared to known strings of characters or text. If a match is not found, the string of characters is separated into sub-strings. The sub-string can be of any length. In one embodiment, the sub-strings are typically three characters long and are referred to as 3-grams (Applicants' Specification paragraph 0009). The applied reference Damashek teaches a pattern recognition technique based on n-gram comparisons among documents that are similar in language and/or topic in that they tend to contain many of the same n-gram. Further, Damashek teaches the reference documents are parsed into n-grams, all the unique n-grams that occur in that reference document (where n is typically fixed at some value that is useful, such as n=5) (col. 3, lines 22-25 and col. 4, lines 24-27; col. 5, lines 24-30). Damashek teaches modifying the query string using a predetermined set of heuristics (col. 4, lines 64 – col. 5, line 5 and col. 8, line 64 – col. 9, line 2). Ortega, on the other hand, teaches a method for predicting the correct spellings of search terms within multiple-term search queries. The prior art evaluates whether any of these related terms has a similar spelling to the misspelled search terms. The spell correction process enters into a loop in which the

spellings of the non-matching terms and the related terms are compared. In each pass of this loop, the process compares a non-matching term to the list of related terms one-by-one. The comparisons are performed using an anagram-type function which compares two characters strings and returns a numerical similarity score (paragraphs 0057 and 0021). Based on the above, Examiner submits that both Demashek and Ortega teach subject matters related to the Applicants' invention and that combining the two references would arrive at Applicants' limitations as claimed.

Applicants argue that neither of the cited references, taken alone or in any combination, teaches or describes such a technique employing sub-string of varying lengths. Damashek describes the use of a fixed length of all n-grams, which specifically states that "n is typically fixed at some value that is useful". In response to the preceding arguments, Examiner respectfully submits that the newly added limitation "*...the sub-strings having varying lengths*" is equivalent to Damashek's n-grams as the Applicants' specification paragraphs 27-31 discloses sub-strings are formed from a series of characters in a given string of characters by extending the sub-strings (i.e., forming variable length sub-strings) based on the frequency of occurrence of the extended sub-strings. The system looks for extensions of frequently appearing sub-string formed by adding one character. For example, the system then looks for frequently appearing 3-grams, 4-grams, or n-grams, where n is any positive integer value greater than 2. Damashek teaches a pattern recognition technique based on n-gram comparisons among documents that are similar in language and/or topic look alike

in that they tend to contain many of the same n-gram. Further, Damashek teaches the reference documents are parsed into n-grams, all the unique n-grams that occur in that reference document (where n is typically fixed at some value that is useful, such as n=5) (col. 3, lines 22-25 and col. 4, lines 24-27; col. 5, lines 24-30). Hence, Damashek's teaching is in conformity with Applicants' limitation of "varying length of sub-strings".

Applicants further argue that Ortega does not even discuss the use of n-grams. Rather, Ortega merely describes a technique for predicting the correct spelling of search terms within multi-term search queried based on previously determined relationships between correctly-spelled and incorrectly-spelled search terms. In response to the preceding arguments, Examiner respectfully submits that Ortega does not have to teach the use of n-grams as the reference was brought in to compliment the limitation "*performing a character-by-character comparison*" which Damashek does not explicitly suggest. Applicants have made a piecemeal analysis of the references. Applicants are therefore reminded that one cannot show nonobviousness by attacking references individually where the rejections are based on combinations of references. See *In re Keller*, 642 F.2d 413, 208 USPQ 871 (CCPA 1981); *In re Merck & Co.*, 800 F.2d 1091, 231 USPQ 375 (Fed. Cir. 1986). Hence, Applicants' attack of Damashek and Ortega references individually cannot be relied upon to show non-obviousness.

THIS ACTION IS MADE FINAL. Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

Art Unit: 2167

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the mailing date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Leslie Wong whose telephone number is (571) 272-4120. The examiner can normally be reached on Monday to Friday 9:30am - 6:30 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John E Breene can be reached on (571) 272-4107. The fax phone number for the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).



Leslie Wong
Patent Examiner
Art Unit 2167

LW
January 10, 2005


Leslie S. Wasserman
Primary Examiner